

--	--	--	--	--	--	--	--	--	--

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 3, 2018/2019

TIB2341 – INTRODUCTION TO BIOINFORMATICS

(All sections / Groups)

31 May 2019
9:00 to 11:00 AM
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use the calculator.

QUESTION 1 [10 MARKS]

- a) Describe briefly how genetic information flows from a DNA sequence to a protein product inside cells during general transfer. [3 marks]
- b) How does nucleotide base pairing forms? Give TWO examples of base pairing. [2 marks]
- c) You performed sequence alignment and found two protein sequences showing high similarity. What may the result tell you about the relationship of these two proteins? [1 mark]
- d) What is transcriptome? [1 mark]
- e) The biological data can be assessed through web based and command-line approaches. Why is the command line software more suitable for the bioinformatics data analysis? [2 marks]
- f) What is the relationship between RefSeq and GenBank? [1 mark]

QUESTION 2 [10 MARKS]

- a) Name TWO bioinformatics applications for the biomolecule structure analysis. [1 mark]
- b) How is sequence alignment applied for evolutionary analysis? [2 marks]
- c) Using the transition probabilities of CpG (P^+) island and non-CpG (P^-) island chains given below; determine whether the sequence **ATTGCGC** is a CpG island. [4 marks]

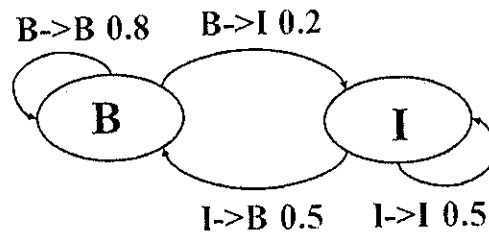
$$P^+ = \begin{bmatrix} & A & C & G & T \\ A & 0.18 & 0.27 & 0.43 & 0.12 \\ C & 0.17 & 0.37 & 0.27 & 0.19 \\ G & 0.16 & 0.34 & 0.37 & 0.13 \\ T & 0.08 & 0.36 & 0.38 & 0.18 \end{bmatrix} \quad P^- = \begin{bmatrix} & A & C & G & T \\ A & 0.30 & 0.20 & 0.29 & 0.21 \\ C & 0.32 & 0.30 & 0.08 & 0.30 \\ G & 0.25 & 0.25 & 0.29 & 0.21 \\ T & 0.18 & 0.24 & 0.29 & 0.29 \end{bmatrix}$$

- d) Define Artificial Neural Networks (ANN). Draw a simple diagram of a 2-layer ANN (input layer, hidden layer, and output). [3 marks]

Continued...

QUESTION 3 [10 MARKS]

- a) Given the following Hidden Markov Model, find the probability of the nucleotide sequence ATT using the forward algorithm (initial $\alpha_B = 0.5$ $\alpha_I = 0.5$) [5 marks]



$$\begin{aligned}
 P(A|B) &= 0.25 & P(T|B) &= 0.40 & P(C|B) &= 0.10 & P(G|B) &= 0.25 \\
 P(A|I) &= 0.25 & P(T|I) &= 0.25 & P(C|I) &= 0.25 & P(G|I) &= 0.25
 \end{aligned}$$

- b) Given the following 2 nucleotide sequences, find their best alignment using dynamic programming and trace back, using Smith-Waterman algorithm (local). [5 marks]

Seq1: GCATCATAACCG

Seq2: TGCATGACG

Match = 2, mismatch = 0, gap = -2

QUESTION 4 [10 MARKS]

- a) Given the following distance matrix, construct the corresponding phylogenetic tree using the distance method (UPGMA). [6 marks]

	A	B	C	D	E	F
A	0					
B	5	0				
C	4	7	0			
D	7	10	7	0		
E	6	9	6	5	0	
F	8	11	8	9	8	0

- b) What are the TWO advantages of distance based methods over other phylogenetic tree construction methods? [2 marks]
- c) Explain ONE difference for each ultrametric and additive distance method in phylogenetic tree. Give one example for each method. [2 marks]

Continued...

QUESTION 5 [10 MARKS]

- a) What is fold recognition? [2 marks]
- b) Why protein threading is the alternative for protein structure prediction apart than sequence similarity search? [2 marks]
- c) Why nondeterministic polynomial problem exists in *ab initio* prediction? [2 marks]
- d) What is the largest protein structure database? [1 mark]
- e) Give one main difference for homology modeling, threading and *ab initio* methods in protein structure predictions? [3 marks]

END OF PAPER